

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>G11B</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 98/35344</b> <b>(43) International Publication Date:</b> 13 August 1998 (13.08.98)
<b>(21) International Application Number:</b> PCT/US98/02740 <b>(22) International Filing Date:</b> 11 February 1998 (11.02.98)  <b>(30) Priority Data:</b> 60/037,911 12 February 1997 (12.02.97) US  <b>(71) Applicant:</b> HYUNDAI ELECTRONICS AMERICA, INC. [US/US]; 3101 North First Street, San Jose, CA 95134 (US).  <b>(72) Inventor:</b> LEE, Jong, Seuk; Hankok Apt. 104, Pungdukchungri Sujiemp, Joninski, Kgounggido (KR).  <b>(74) Agents:</b> LANG, Dan, H. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).		<b>(81) Designated States:</b> DE, GB, JP, KR.  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> A NONVOLATILE MEMORY STRUCTURE  <b>(57) Abstract</b>  The present invention provides a novel nonvolatile Flash EEPROM array design which allows for array, block or sector erase capabilities. The relatively simple transistor design layout of the present invention allows small portions of the EEPROM array to be erased without affecting data stored in the remaining portion of the array. In addition, given the block structured layout of the Flash EEPROM array, adjacent blocks in the array can share transistor control circuitry, thus minimizing the size of the array. The novel nonvolatile Flash EEPROM array preferably comprises a plurality of blocks which comprise a plurality of sectors of NOR-gate transistors. Each transistor has a drain, a source, and a control gate. Preferably, the drains of each transistor in a column are electrically coupled, the control gates of each transistor in a row are electrically coupled, and the sources of all the transistors in a sector are electrically coupled. A sector of the nonvolatile Flash EEPROM array preferably comprises 8 rows and 512 columns of transistors and a block preferably comprises 128 vertically stacked sectors.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## A NONVOLATILE MEMORY STRUCTURE

5

### CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application No. 60/037,911 filed February 12, 1997, incorporated herein by reference.

10

### BACKGROUND OF THE INVENTION

The present invention relates generally to nonvolatile memory devices and more specifically to a Flash EEPROM memory design utilizing a novel NOR-gate transistor array architecture.

15

Most computers currently use magnetic disk drives for data storage. However, disk drives tend to be bulky and have a number of moving parts. Consequently, they are prone to reliability problems and consume a significant amount of power. Moreover, as PC's and other digital devices such as cameras and PDA's become smaller and smaller, magnetic disk drive storage becomes even more impractical.

20

Recently, Flash electrically erasable programmable read-only memory (EEPROM) has emerged as a new source of nonvolatile memory storage. Flash EEPROM memory devices typically comprise an array of floating gate transistors for storing data in digital form.

25

FIG. 1 illustrates the cross-section of a typical NMOS floating gate transistor cell 100 used in Flash EEPROM devices. NMOS transistor cell 100 typically comprises a p-type substrate 102 having a drain region 104 and a source region 106. Drain region 104 and source region 106 typically consist of N+ diffusion regions within p-type substrate 102. A channel region 108 in substrate 102 separates drain region 104 from source region 108.

30

35

Positioned above substrate 102 and drain and source regions 104, 106 are a floating gate 110 and a control gate 112, both which typically are formed of polysilicon. Floating

gate 110 is separated from substrate 102 by a thin dielectric layer 114, which in most cases, comprises silicon dioxide. Similarly, a dielectric layer 116 separates floating gate 110 and control gate 112. The entire structure is overlaid by an oxide insulating layer 118, and means are provided for applying a source voltage  $V_s$  through oxide layer 118 to source region 106, a gate voltage  $V_g$  to control gate 112, and a drain voltage  $V_d$  through oxide layer 118 to drain region 104.

To program Flash EEPROM transistor cell 100, drain 104 and control gate 112 are raised to voltage potentials above the voltage potential of source 106. For example, drain 104 is raised to a potential  $V_d$  of about 5 volts and control gate 112 is raised to a potential  $V_g$  of about 12 volts. Source 106 is typically grounded. As illustrated in FIG. 1, under such conditions, the current generates hot electrons which become trapped in floating gate 110. This electron injection increases the floating gate threshold by about 3 to 5 volts.

To erase the Flash EEPROM transistor cell 100, drain 104 is typically floated, control gate 112 is grounded and a voltage of about 9 to 12 volts is applied to source 106 for a few milliseconds. As a result, the electrons stored on the floating gate 110 will tunnel through dielectric 114 to drain 104.

Finally, to read cell 100 (i.e., to determine whether a one or a zero is stored in the cell), source 106 is typically held at ground potential and a voltage of about 5 volts is applied to control gate 112. A potential of about 1 to 2 volts is applied to drain 104. Under these conditions, an unprogrammed cell (i.e., no electrons on the floating gate) conducts a current of about 25 to 50 microamps. A programmed cell does not conduct.

As illustrated in FIG. 2, a typical Flash EEPROM array 200 comprises a plurality of transistors 202 arranged in rows and columns. In accordance with this well known arrangement, the drains D of each cell 202 in a column are connected to a common bit line 204. Similarly, the control gates of each cell 202 in each row are connected to a common word line 206.

The source lines of all the cells 202 in the entire array are tied to a common source line 208.

With this configuration, cells 202 of array 200 may be individually programmed, but all the cells in array 200 are  
5 erased simultaneously because the sources of all the cells are tied together. Thus, to re-program any portion of the array, the entire array first must be erased and then re-programmed. If some of the information stored in the array is to remain  
10 the same, that information must be saved in memory during the erase process and then re-programmed back into the array. As one skilled in the art can appreciate, a complex control and memory system is needed to program this type of Flash EEPROM array.

Various different Flash EEPROM array designs have been  
15 developed to overcome some of the problems associated with having to erase an entire memory array before re-programming it. For example, NAND-gate transistor Flash EEPROM designs have been developed to overcome some of the large block erasing problems associated with the NOR-gate designs. See,  
20 for example, "An Experimental 4-Mbit CMOS EEPROM with a NAND-Structured Cell," M. Momodomi, et al., IEEE Journal of Solid-State Circuits, vol. 24, No. 5, Oct. 1989, pp. 1238-1243, and  
25 "A High-Density NAND EEPROM with Block-Page Programming for Microcomputer Applications," Y. Iwata, et al., IEEE Journal of Solid-State Circuits, vol. 25, No. 2, April 1990, pp. 417-424. However, while these NAND-gate transistor designs allow for smaller sector sizes and, thus, smaller block erasures, they also tend to have more overhead, causing much slower processing speeds.

30 Thus, a NOR-gate Flash EEPROM design which permits easy erasing and programming control of smaller transistor sectors and blocks is needed which overcomes the shortcomings of the prior art.

## SUMMARY OF THE INVENTION

The present invention provides a novel nonvolatile Flash EEPROM array design which allows for array, block or sector erase capabilities. The relatively simple transistor design layout of the present invention allows small portions of the EEPROM array to be erased without affecting data stored in the remaining portions of the array. In addition, given the block structured layout of the Flash EEPROM array of the present invention, adjacent blocks in the array can share transistor control circuitry, thus minimizing the size of the array.

According to one aspect of the present invention, the novel nonvolatile Flash EEPROM array comprises a NOR-gate transistor matrix having at least two rows and two columns of transistors. Each transistor has a drain, a source, and a control gate. In accordance with a preferred embodiment of the invention, the drains of each transistor in a column are electrically coupled, the control gates of each transistor in a row are electrically coupled, and the sources of all the transistors in the matrix are electrically coupled.

In accordance with another aspect of the present invention, a sector of the nonvolatile Flash EEPROM array preferably comprises 8 rows and 512 columns of transistors in which the drains in each column of the sector are electrically coupled, the control gates in each row of the sector are electrically coupled and all the sources within the sector are electrically coupled. In accordance with this aspect of the invention, a sector preferably comprises 512 bytes of nonvolatile memory.

In accordance with yet another aspect of the present invention, a block of the nonvolatile Flash EEPROM array preferably comprises a plurality of vertically stacked sectors. In accordance with this aspect of the invention, the drains of all transistors within a column of the block are electrically coupled with a common bit line. That is, the drains in a column of one sector are electrically coupled to the drains in the corresponding column of adjacent sectors. In this manner, a block comprises a plurality of continuous columns.

Also, as with the sector design of the present invention, the control gates in each row of the block are electrically coupled. However, in accordance with a preferred embodiment of the invention, only the sources within each sector are  
5 coupled together. The common sources in each sector are not tied to the common sources of other sectors in the block.

In accordance with yet another aspect of the present invention, a block preferably comprises 128 vertically stacked sectors. In accordance with this aspect of the invention, a  
10 block preferably comprises 64 Kbytes of nonvolatile memory.

In accordance with still another aspect of the present invention, the nonvolatile Flash EEPROM array may comprise a plurality of rows and/or columns of blocks, and adjacent blocks can share control circuitry.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a side cross-sectional view of a typical NMOS floating gate transistor;

FIG. 2 is a schematic diagram of a prior art Flash EEPROM  
20 transistor array configuration;

FIG. 3 is a schematic diagram of the transistor configuration of one sector of the Flash EEPROM design of the present invention;

FIG. 4 is a top plan view of a circuit layout of a  
25 portion of the sector shown in FIG. 3;

FIG. 5 is a schematic diagram of the transistor configuration of one block of the Flash EEPROM design of the present invention;

FIG. 6 is a block diagram of the Flash EEPROM planar  
30 array of the present invention;

FIG. 7 is a circuit diagram of a local decoder used in the embodiment of the Flash EEPROM planar array illustrated in FIG 6.

## DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention provides a novel nonvolatile Flash EEPROM array design which allows for array, block or sector erase capabilities. Accordingly, a portion (i.e., a block or sector) of the memory array of the present invention can be erased and re-programmed with new data without affecting the entire array. While the preferred embodiment of the present invention is described herein with reference to a specific number of transistors in a sector and a specific number of sectors in a block, one skilled in the art will appreciate that any number of transistor may make-up a sector, and any number of sectors may make-up a block. Moreover, any number of blocks may be combined to form the entire Flash EEPROM array. Accordingly, the present invention is not limited to the specific embodiments disclosed herein.

In the figures, similar components and/or features have the same reference label. The various components are distinguished by following the reference label by a dash and a second label that distinguishes among the similar components. If only the first reference label is used, the description is applicable to any one of the several similar components.

Referring now to FIG. 3, a sector of Flash EEPROM memory 300 preferably comprises a plurality of MOS NOR-gate transistor cells 302 arranged in rows and columns. In accordance with a preferred embodiment, sector 300 comprises 8 rows and 512 columns of transistor cells 302. Thus, in accordance with this aspect of the present invention, sector 300 is configured to store 512 bytes of data; each column being 8 bits or 1 byte of data.

Each transistor cell 302 comprises a drain 304, a source 306, a control gate 308, and a floating gate 310. As illustrated in FIG. 3, drains 304 of each cell 302 in a column are connected via a bit line 312. Similarly, control gates 308 of each cell 302 in a row are connected via a word line 314. Sources 306 of each cell 302 in sector 300 are connected to a single source connection 316.

FIG. 4 illustrates the layout of transistor cells 302 of a portion of sector 300. As shown in FIG. 4, transistor cells



302 in a column (i.e., along bit line 312) may share common source wells 306 and common drain wells 304. For example, in accordance with the illustrated embodiment, transistor cell 302-1 shares a common source well 306 with transistor cell 302-2, and transistor cell 302-2 shares a common drain well 304 with cell 302-3. Similarly, transistor cell 302-3 shares a common source well 306 with transistor cell 302-4, and so on. This configuration allows the transistors to be more densely packed in the array, thus, permitting greater storage capacity in an area smaller than other NOR-gate Flash EEPROM devices current known in the art.

As mentioned previously with reference to FIG. 3, drains 304 of cells 302 in a column are connected via bit line 312. As illustrated more clearly in FIG. 4, bit lines 312 preferably comprise first metal lines 402 connected to drains 304 by metal to N+ diffusion vias 404. In accordance with this aspect of the invention, by connecting all cells 302 in a column with bit line 312, and transistor cells 302 along the bit line may be controlled by a common bit line driver, column selector, and sense amplifier, thus reducing the analog circuitry overhead necessary for implementing the memory device.

Control gates 308 of transistor cells 302 in a row are all connected via word line 314, which, in accordance with a preferred embodiment, is a polysilicon layer 406. As one skilled in the art will appreciate, polysilicon layer 406 forms the actual control gates 308 of the transistor cells, as well as the connections between the adjacent gates 308. Floating gates 310 of transistor cells 302 also preferably are formed of polysilicon, but are not connected between the individual transistor cells. The polysilicon floating gates 310 typically reside below polysilicon layer 406 but, for clarity, are not illustrated in FIG. 4.

As mentioned previously, sources 306 of each transistor cell 302 in a sector are tied to a common source connection 316. In accordance with a preferred embodiment of the present invention, all sources 306 of transistor cells 302 in a common row are coupled together with an N+ diffusion layer/line 408.

As one skilled in the art will appreciate, N+ diffusion layer 408 forms the source wells 306 of each transistor cell 302, as well as connects source wells 306 along a row.

In order for all sources 306 within sector 300 to be tied to single source connection 316, N+ diffusion lines 408 preferably are tied together within sector 300. As illustrated in FIG. 4, one or more second metal lines 414 run perpendicular to N+ diffusion lines 408 and are connected to N+ diffusion lines 408 with metal to N+ diffusion vias 416. In this manner, second metal lines 414 interconnect all the parallel N+ diffusion lines 408 in sector 300, effectively linking all sources 306 in sector 300 to common source connection 316. As one skilled in the art will appreciate, second metal lines 414 may reside either at the same fabrication layer or at a different fabrication layer as first metal lines 402. However, in accordance with a preferred embodiment of the present invention, second metal lines 414 reside at the same layer as first metal lines 402.

In addition, one or more third metal lines 410 preferably run parallel to N+ diffusion lines 408 and connect to second metal lines 414 with metal to metal vias 412. By connecting third metal lines 410 to N+ diffusion lines 408 via second metal lines 414, the overall resistance of N+ diffusion lines 408 are effectively reduced, thus reducing the power requirements of the Flash EEPROM array. In accordance with a preferred embodiment of the invention, third metal lines 410 are located at different layer from first and second metal lines 402, 414.

While the illustrated embodiment shows second metal lines 414 positioned between about every third bit line 312, one skilled in the art will appreciate that second metal lines 414 may have any suitable spacing in sector 300. For example, first metal lines can be positioned between about every 16 lines 312 or every 32 lines 312, depending on the resistance of the N+ lines between one contact (416) and another contact (416). Alternatively, only one second metal line 414 may be utilized. Similarly, while the illustrated embodiment shows third metal lines 410 positioned over about every other row

of drains 304, one should appreciate that any number of third metal lines 410 may be used. For example, third metal lines 410 may be positioned directly over and coupled to N<sup>+</sup> diffusion lines 408 with metal to N<sup>+</sup> diffusion vias.

5 Alternatively, only one third metal line 410 may be utilized. Accordingly, the present invention is not limited to the illustrated embodiment.

Referring now to FIG. 5, a memory block 500 of the Flash EEPROM array of the present invention is illustrated. In particular, block 500 preferably comprises a plurality of sectors 300 stacked vertically one on top of the next. All drains 304 of transistor cells 302 in a column of block 500 are connected to a common bit line 312 via first metal lines 402 (see FIG. 4). In this manner, the vertically stacked sectors 300 all share common bit lines 312. Also, just as transistor cells 302 within a column of a sector can share common source 306 and drain 304 wells (see FIG. 4), transistor cells 302 in adjacent sectors 300 within block 500 can also share common source 306 and drain 304 wells. As mentioned previously, this configuration permits the transistors to be more densely packed within the array. Also, by sharing common bit lines between sectors 300, the total number of bit line drivers, sense amplifiers and column selectors are reduced.

In accordance with a preferred embodiment of the invention, block 500 preferably comprises 128 stacked sectors 300. Accordingly, each block 500 comprises 1024 word lines 314 (8 per sector), 128 source connections 316 (one for each sector), and 512 bit lines 312. Thus, block 500 can store up to 64 Kbytes of information:

1 sector = 512 columns and 8 rows of transistors =  
512 bytes of storage  
1 block = 128 sectors x 512 bytes = 64 Kbytes of storage

Referring now to FIG. 6, a preferred embodiment of a nonvolatile Flash EEPROM planar array 600 of the present invention is illustrated. Planar array 600 preferably comprises a plurality of blocks 500 aligned in a one or two dimensional configuration. In accordance with the illustrated

embodiment, two rows 602 and four columns 604 of blocks 500 are shown.

To control the operation of planar array 600, word line decoders 606, source decoders 608, global decoders 610, and bit line decoders 612 are used. As illustrated in FIG. 6, adjacent horizontal blocks 500 share word line decoders 606 and source decoders 608. For example, in row 602-1 of FIG. 6, block 500-1-1 shares word line decoder 606-1-1 with block 500-1-2, and block 500-1-3 shares word line decoder 606-1-2 with block 5-1-4. Similarly, blocks 500-1-2 and 500-1-3 share source decoder 608-1-2. The second row 602-2 of planar array 600 is similarly configured.

Also, in accordance with a preferred embodiment of the present invention, planar array 600 is configured so that vertically adjacent blocks 500 share bit line drivers, column selectors and sense amplifiers. For clarity purposes bit line drivers, column selectors and sense amplifiers have been grouped together as bit line decoders 612. However, one skilled in the art will appreciate that these components may comprise separate circuitry and they need not be group as one component.

In reading, writing and erasing planar array 600 or any block 500 or sector 300 therein, a controller (not shown) typically sends control signals to the various decoders 606-612, directing the decoders to apply suitable voltages to the various bit lines 312, word lines 314 and source connections 316 of the transistors cells 302 being read, written to, or erased. For example, to read one or more transistor cells 302 in planar array 600, the controller sends the appropriate signals to the specific word line decoder(s) 606, source decoder(s) 608, global decoder(s) 610, and bit line decoder(s) 612 which control the particular transistor cells 302 being read. The various decoders 606-612 then control the voltages on the particular word lines 314, bit lines 312, and common source connections 316 connected to the particular transistor cells 302 being read. Similar operations occur for the write and erase processes.

To read a transistor cells 302 in planar array 600, a voltage in the range of about 4 to about 7 volts and preferably about 5 volts is applied to control gate 308 of the transistor cell being read via word line 314. Similarly, source 306 preferably are grounded and the voltage on drain 304 is measured using bit line decoder 612. In accordance with this aspect of the invention, transistor cell 302 is programmed with a zero (0) if the voltage on drain 304 is in the range of about .5 to about 2 volts and preferably about 1 volt. Transistor cell 302 contains a one (1) if no voltage is detected on drain 304.

In accordance with another embodiment of the present invention, instead of detecting the voltage on drain 304 to read transistor cell 302, a voltage of about .5 to about 2 volts and preferably about 1 volt is applied to drain 304 via bit line 312, and the current flow in transistor cell 302 is detected using bit line decoder 612. An unprogrammed cell typically conducts about 25 to 50 microamps. A programmed cell does not conduct.

Similarly, to program a transistor cell 302 in EEPROM array 600, a voltage in the range of about 4 to about 7 volts and preferably about 5 volts is applied to control gate 308 of the transistor cell being programmed via word line 314. In addition, source 306 of the transistor cell preferably is grounded and a voltage in the range of about 5 to about 9 volts and preferably about 6 volts is applied to drain 304. Under these conditions, electrons from the P-substrate of the transistor cell tunnel through an oxide layer to floating gate 310, thus "programming" a zero (0) therein.

Finally, in accordance with the present invention, the erase function preferably is performed on one or more sectors 300 or one or more blocks 500 of EEPROM array 600. Thus, to erase one or more sectors 300, all control gates 308 in sector(s) 300 preferably are grounded, all drains 304 are allowed to float, and a voltage in the range of about 8 to about 12 volts and preferably about 9.5 volts is applied to all sources 306 via common source connection 316. Under these conditions, electrons that were placed in floating gates 310

during the programming process tunnel back through the oxide insulating layer and into drains 304. In accordance with this aspect of the invention, transistor cells 302 within the sector(s) 300 being erased which were initially programmed with a zero (0) are "erased" back to a one (1) state (i.e., no voltage is detected on drain 304 during the read process.

Given the particular configuration of the EEPROM array of the present invention, one or more transistor cells 302 can be read or programmed at any particular time. However, to erase a transistor cell in a particular sector, the entire sector must be erased. Accordingly, each source decoder 608 is configured to control the voltage potentials on each one of the 128 common source connections 316 in block 500. In accordance with this aspect of the invention, during the erase process, source decoder 608 applies a suitable voltage to or "selects" one or more entire sectors at a time. If source decoder 608 is shared by two or more blocks 500, source decoder 608 preferably is configured to control the sectors in the blocks separately. For example, referring to FIG. 6, source decoder 608-1-2 can select a specific sector 300 in block 500-1-2 and a different sector 300 in block 500-1-3 simultaneously.

Similarly, each bit line decoder 612 is configured to control each one of the 512 bit lines 312 in block 500 separately, and each word line decoder 606 is configured to control each one of the 1024 word lines 314 in block 500 separately. Thus, specific transistor cells 302 within a particular sector can be read or programmed by applying suitable voltages to specific bit lines 312 and word line 314 within the sector. For example, to program one or more transistor cells 302 within a sector of block 500-1-1, suitable voltages are applied to the drains 304 and control gates 308 the particular transistor cells 302 by bit line decoder 612-1 and word line decoder 606-1-1 respectively.

Referring now to FIG. 7, the configuration and operation of word line decoder 606 will be discussed. Specifically, word line decoder 606 is electrically coupled to a global decoder 610 and comprises a partial decoder 702 and a

plurality of AND-gates 704. Global decoder 610 receives control signals from a controller and generates a plurality of outputs 706 which act as inputs to AND-gates 704. In accordance with a preferred embodiment of the present invention, global decoder 610 includes 128 outputs 706, one for each sector 300 in block 500.

Similarly, partial decoder 702 receives control signals from the controller and generates a plurality of outputs 708 (preferably 8) which also act as inputs to AND-gates 704. The combination of input signals from global decoder 610 and partial decoder 702 generates signals which control one or more word lines 314 in block 500 (See FIGS. 3-5). In accordance with this aspect of the invention, the output of word line decoder 606 is a plurality of word line connections 710 which are configured to electrically couple to word lines 314. In accordance with a preferred embodiment, word line decoder 606 comprises 1024 word line connections 710; one for each word line 314 in block 500.

For example, for word line decoder 606 to apply a voltage to or "select" the word line WL12 (see FIG. 7), global decoder 610 generates a signal on output 706-2 pursuant to a control signal from the controller. Similarly, partial decoder 702 generates a signal on partial decoder output 708-4. AND-gate 704 combines the two signals and generates a signal on word line connection 710-12 which is coupled to word line WL12. As one skilled in the art will appreciate, a particular AND-gate 704 will not generate an output unless both the global decoder 606 and the partial decoder 702 transmit high voltages to the particular AND-gate.

To select all word lines 314 within a sector, global decoder 606 generates a high voltage signal for the particular sector selected, and all eight partial decoder outputs 708 are sent high. In accordance with this aspect of the invention, the combination of the output(s) from global decoder 606 with the high outputs from partial decoder 702 produces high voltage signals on all word line connection 710 in the selected sector.

In accordance with another aspect of the present invention, the controller for the nonvolatile Flash EEPROM array of the present invention includes circuitry configured to receive externally generated control signals and convert the external signals to internal Flash EEPROM control and data signals. For example, the external signals may be particular address and data signals generated by an Address + Data Pin Interface Protocol. In accordance with this aspect of the present invention, the external address and data signals may be multiplexed into one external signal which can be processed by the control circuitry of the EEPROM array. In addition, the Flash EEPROM control circuitry may be configured to interface with and support MSDOS, MSDOS Disk Internal Operations, ATA signal protocol, PCMCIA/ATA signal protocol, Compact Flash signal protocol, and any other memory access protocols currently known or developed in the future. For a more detailed discussion of these well known protocols and how they interface with EEPROM or Flash EEPROM devices, see for example, AIMS Specification Release 1.01 and PC Card ATA Mass Storage Specification Release 1.02, both of which are incorporated herein by reference.

In accordance with yet a further aspect of the present invention, the control circuitry may comprise an error control circuit implementing a fault tolerance scheme such as a 1-bit error correcting Hamming Code to support fault tolerance in a multiple bit per transistor coding scheme. As one skilled in the art will appreciate, such error control circuits are well known in the art.

In accordance with yet another aspect of the present invention, the Flash EEPROM control circuitry further comprises a cache memory to optimize both read and write operations. For example, to perform a fast read operation, data stored in the flash memory cell can be dumped into the cache memory, and the data subsequently can be read serially from the cache memory. Similarly, to perform a fast write operation, the data to be written to the flash memory cell can be dumped into the cache memory, and the write operation can occur from the cache memory to the flash memory cell.



In accordance with yet a further aspect of the present invention, all the Flash EEPROM control circuitry is fabricated into a single IC chip. In accordance with one embodiment of the invention, the single IC chip may include the cache memory device.

In addition, in accordance with another embodiment of the invention, the nonvolatile Flash EEPROM transistor array of the present invention may be fabricated into the same IC chip as the control circuitry and cache memory. In accordance with this aspect of the invention, one IC chip will include everything necessary to implement the Flash EEPROM array of the present invention.

In conclusion, the present invention comprises a novel EEPROM design configured for implementing nonvolatile memory structures into large arrays which can be used to form hierarchical memory organizations. While the above is a complete description of the preferred embodiments of the invention, various alternatives, modifications, and equivalents may be used. For example, the size of each array, block and/or sector may be altered to achieve specific design considerations. In addition, a different control circuitry design may be utilized to implement the nonvolatile memory array of the present invention. Therefore, the above description should not be taken as limiting the scope of the invention which is defined by the appended claims.

WHAT IS CLAIMED IS:

1           1.    A nonvolatile NOR-gate transistor architecture  
2 comprising:

3           a transistor matrix comprising a plurality of MOS  
4 transistors disposed in a matrix of at least two rows and at  
5 least two columns, each transistor having a source, a control  
6 gate, and a drain, said control gates in each row being  
7 electrically coupled, said drains in each column being  
8 electrically coupled, and all of said sources being  
9 electrically coupled.

1           2.    The nonvolatile transistor architecture of  
2 claim 1 wherein said sources disposed in a row are  
3 electrically coupled via N+ diffusion.

1           3.    The nonvolatile transistor architecture of  
2 claim 1 wherein said drains are electrically coupled by a  
3 first metal generally disposed at a first layer of said  
4 architecture.

1           4.    The nonvolatile transistor architecture of  
2 claim 2 wherein said drains are electrically coupled by a  
3 first metal generally disposed at a first layer of said  
4 architecture and said rows of said sources are electrically  
5 coupled by a second metal generally disposed at the same layer  
6 of said architecture as said first layer.

1           5.    The nonvolatile transistor architecture of  
2 claim 2 wherein said drains are electrically coupled by a  
3 first metal generally disposed at a first layer of said  
4 architecture and said rows of said sources are electrically  
5 coupled by a second metal generally disposed at a second layer  
6 of said architecture different from said first layer.

1           6.    The nonvolatile transistor architecture of  
2 claim 1 wherein said plurality of MOS transistors comprises  
3 sufficient memory to load MSDOS.

1           7.    The nonvolatile transistor architecture of  
2 claim 1 wherein said plurality of MOS transistors comprises  
3 sufficient memory to load MSDOS and Disk Internal Operations.

1           8.    The nonvolatile transistor architecture of  
2 claim 1 wherein a sector of said transistor architecture  
3 comprises 8 rows and 512 columns of transistors.

1           9.    The nonvolatile transistor architecture of  
2 claim 8 wherein a block of said transistor architecture  
3 comprises a plurality of vertically stacked sectors.

1           10.   The nonvolatile transistor architecture of  
2 claim 9 wherein said block comprises 128 sectors.

1           11.   The nonvolatile transistor architecture of  
2 claim 9 wherein said transistor architecture comprises a  
3 plurality of said blocks disposed in a matrix of at least one  
4 row and at least one column of said blocks.

1           12.   The nonvolatile transistor architecture of  
2 claim 11 further comprising control circuitry including at  
3 least one source decoder, at least one bit line decoder and at  
4 least one word line decoder.

1           13.   The nonvolatile transistor architecture of  
2 claim 12 wherein said at least one source decoder controls  
3 inputs to the sources of at least two adjacent blocks in a  
4 row.

1           14.   The nonvolatile transistor architecture of  
2 claim 12 wherein said at least one word line decoder controls  
3 inputs to the word lines of at least two adjacent blocks in a  
4 row.

1           15.   The nonvolatile transistor architecture of  
2 claim 12 wherein said at least one bit line decoder controls

3 inputs to the bit lines of at least two adjacent blocks in a  
4 column.

1 16. A nonvolatile NOR-gate transistor architecture  
2 comprising:

3 a plurality of sectors each comprising a matrix of MOS  
4 transistors disposed in rows and columns, each transistor  
5 having a source, a control gate, and a drain, said control  
6 gates in each row being electrically coupled, said drains in  
7 each column being electrically coupled, and all of said  
8 sources being electrically coupled.

1 17. The nonvolatile transistor architecture of  
2 claim 16 wherein a first sector is disposed above a second  
3 sector.

1 18. The nonvolatile transistor architecture of  
2 claim 16 wherein a first sector is disposed adjacent a second  
3 sector.

1 19. The nonvolatile transistor architecture of  
2 claim 16 wherein said drains of a first sector and said drains  
3 of a second sector are electrically coupled.

1 20. The nonvolatile transistor architecture of  
2 claim 16 further comprising a controller electrically coupled  
3 to at least two sectors.

1 21. The nonvolatile transistor architecture of  
2 claim 20 wherein said controller permits access to only one  
3 sector at a time for performing a function on said  
4 transistors, said function selected from a group comprising  
5 reading, writing, and erasing.

1 22. The nonvolatile transistor architecture of  
2 claim 20 wherein said controller permits access to at least  
3 two sectors at a time for performing a first function on said  
4 transistors of a first sector and a second function on said

5 transistors of a second sector, said first function and said  
6 second function selected from a group comprising reading,  
7 writing, and erasing.

1 23. The nonvolatile transistor architecture of  
2 claim 20 wherein said controller implements a fault tolerance  
3 scheme to support a multiple bit per transistor coding scheme.

1 24. The nonvolatile transistor architecture of  
2 claim 23 wherein said fault tolerance scheme in a one-bit  
3 error-correcting Hamming code.

1 25. The nonvolatile transistor architecture of  
2 claim 20 further comprising at least one decoder electrically  
3 coupled to said controller and to said sources of at least one  
4 sector, said at least one decoder selecting said sector and  
5 transmitting source signals to said sources in response to  
6 control signals received from said controller.

1 26. The nonvolatile transistor architecture of  
2 claim 25 wherein said control signals are external address and  
3 data signals.

1 27. The nonvolatile transistor architecture of  
2 claim 25 further comprising a cache memory electrically  
3 coupled to said at least one decoder.

1 28. The nonvolatile transistor architecture of  
2 claim 26 wherein a plurality of said external address and data  
3 signals are multiplexed into one control signal.

1 29. The nonvolatile transistor architecture of  
2 claim 25 wherein said sectors, said controllers, and said at  
3 least one decoder are formed as a single integrated circuit.

1 30. The nonvolatile transistor architecture of  
2 claim 20 wherein said controller provides control signals that  
3 support a protocol convention selected from a group comprising

- 4 ATA, PCMCIA/ATA, and Compact Flash signal protocol
- 5 conventions.

1/6

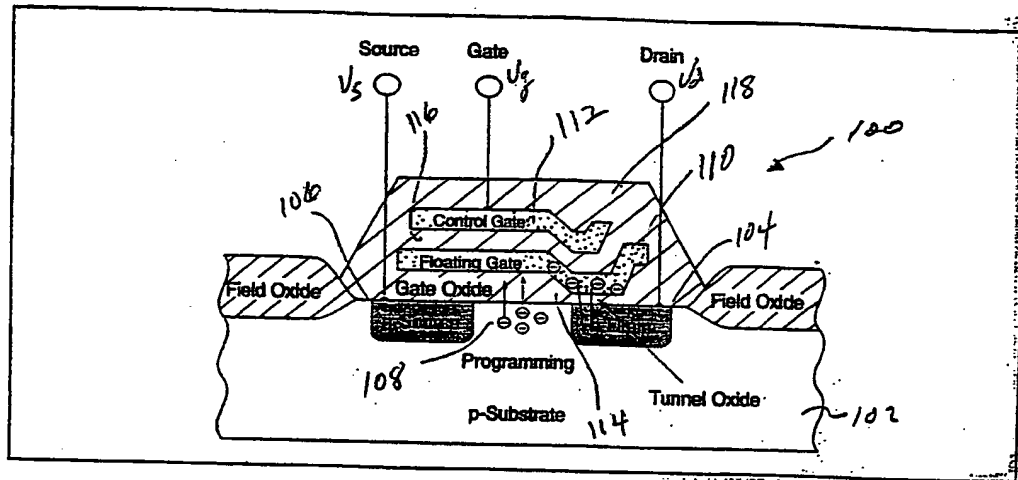


FIG. 1  
(Prior Art)

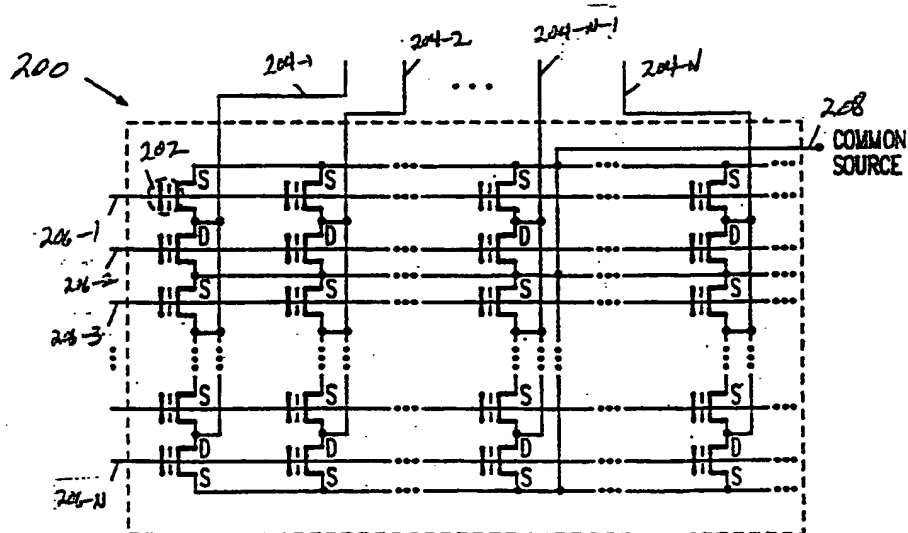


FIG. 2  
(Prior Art)

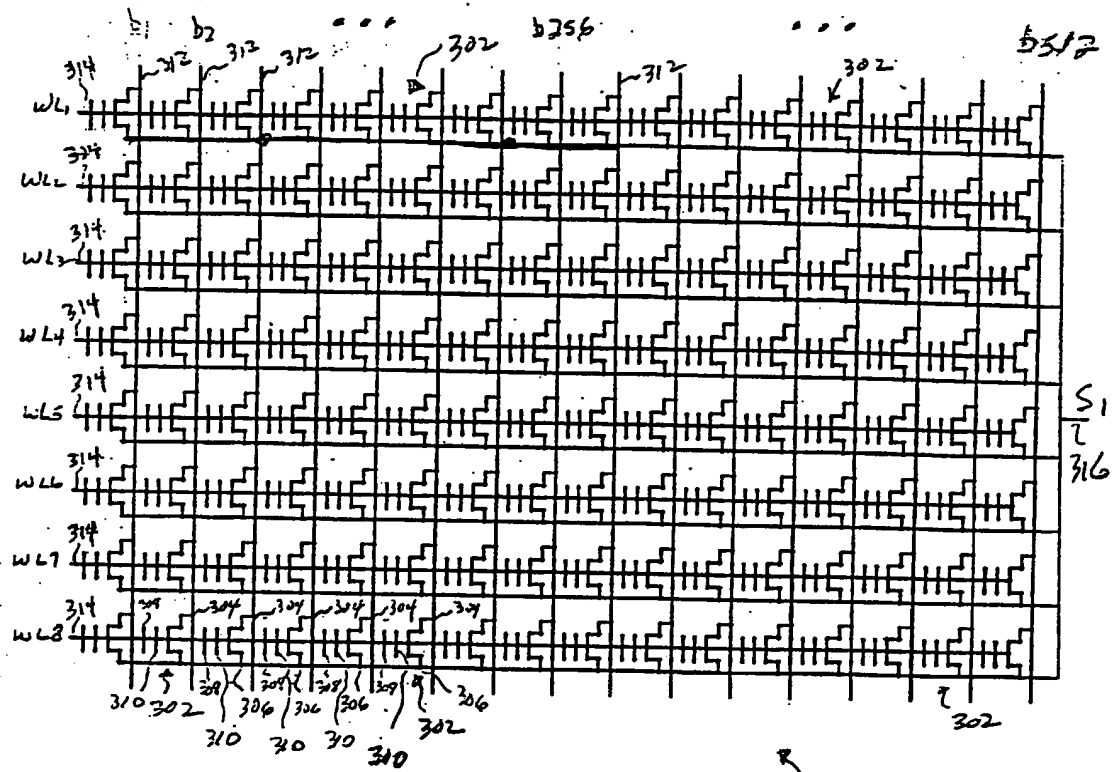


FIG. 3

300





4/6

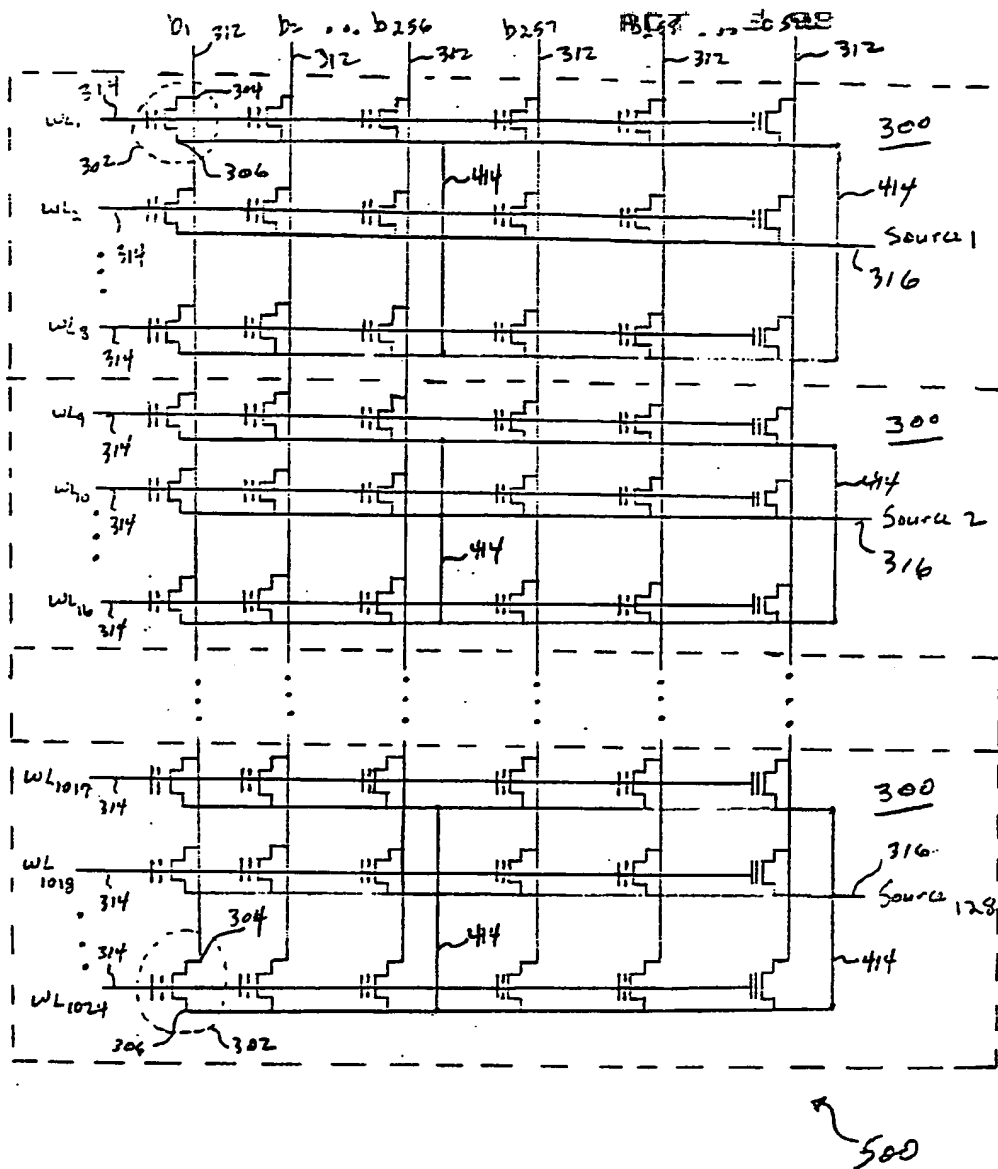


FIG. 5

5/6

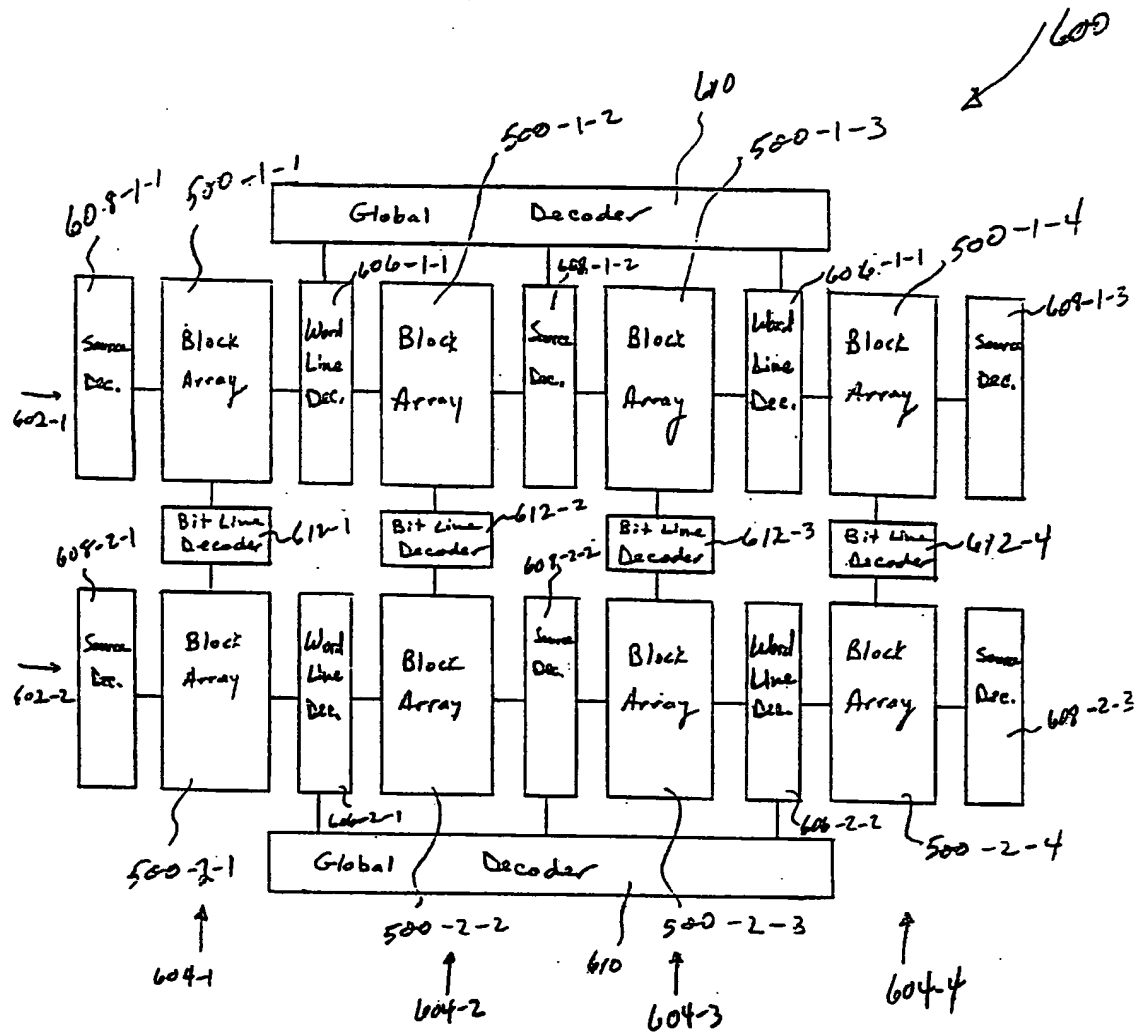


FIG. 6

6/6

